

Journal of Educational Statistics
Winter 1978, Volume 3, Number 4, pp. 305-317

FORGETTING, GUESSING, AND MASTERY:
THE MACREADY AND DAYTON MODELS
REVISITED AND COMPARED WITH A LATENT TRAIT APPROACH

Wim J. van der Linden

Twente University of Technology
Enschede, The Netherlands

*Key words: Mastery Testing, Criterion-Referenced Testing,
Latent Trait Theory, Latent Class Models*

ABSTRACT

Macready and Dayton (1977) introduced two probabilistic models for mastery assessment based on an idealistic all-or-none conception of mastery. Although these models are in statistical respects complete, the question is whether they are a plausible rendering of what happens when an examinee responds to an item. First, a correction is proposed that takes account of the fact that a master who is not able to produce the right answer to an item may guess. The meaning of this correction and its consequences for estimating the model parameters are discussed. Second, Macready and Dayton's latent class models are confronted with the three-parameter logistic model extended with the conception of mastery as a region on a latent variable. It appears that from a latent trait theoretic point of view, the Macready and Dayton models assume item characteristic curves that have the unrealistic form of a step function

with a single step. The implications of the all-or-none conception of mastery for the learning process will be pointed out shortly. Finally, the interpretation of the forgetting parameter of the Macready and Dayton models is discussed and approached from a latent trait theoretic point of view.

INTRODUCTION

The article of Macready and Dayton (1977) proposes two statistically elegant models for the assessment of mastery. In both models mastery is conceived as an all-or-none quality of an examinee with respect to a homogeneous item domain. Mastery implies that the examinee can solve all items of the domain; his or her "true response vector" for the entire domain contains only "ones", indicating that the examinee can respond correctly to all items. An observed failure of a master to succeed on an item is supposed to be the result of extraneous influences that made him or her a victim of forgetting. For nonmastery the opposite holds. An examinee is a nonmaster when he or she does not have the knowledge and skills required for passing any item of the domain; his or her "true response vector" for the entire domain contains only zeros, indicating that he or she will fail all items. An observed success of a nonmaster is attributed to guessing. According to this all-or-none conception of mastery, also called a state model by Meskauskas (1976), there is no state between mastery and nonmastery. Every examinee is either a master or a nonmaster, and it is not possible to have a true score vector containing zeros and ones for the given domain.

Denoting the probabilities of guessing right and forgetting for item i by α_i and β_i , respectively, the probability of responding correctly to item i can, according to Model I of Macready and Dayton, be written as

$$P_i(+) = \begin{cases} 1-\beta_i & \text{for a master} \\ \alpha_i & \text{for a nonmaster.} \end{cases} \quad (1)$$

Assuming local independence among responses and denoting the proportion of examinees who are master and nonmaster by $\bar{\theta}$ and $\bar{\theta}$, respectively, Macready and Dayton derived their Model I, which describes the probability of the j th observed response vector on an n -item test, as

$$P(j) = \left[\prod_{i=1}^n \alpha_i^{a_{ij}} (1-\alpha_i)^{1-a_{ij}} \right] \bar{\theta} + \left[\prod_{i=1}^n \beta_i^{1-a_{ij}} (1-\beta_i)^{a_{ij}} \right] \bar{\theta}, \quad (2)$$

where $a_{ij} = 1$ if the response to the i th item for the j th vector is right, and $a_{ij} = 0$ otherwise. Model II follows upon assuming $\alpha_i = \alpha$ and $\beta_i = \beta$ for all items.

For both models, Macready and Dayton outlined how maximum likelihood estimates of the model parameters can be obtained. Moreover, they showed procedures for testing the fit of the models to actual test data, for establishing optimal rules for mastery decisions, and for determining the number of items minimally sufficient to keep the proportion of misclassified examinees below a previously specified standard. Emrick and Adams (1969), Emrick (1971), and Besel (1973) proposed models that correspond to both models of Macready and Dayton, but that lack, however, these options. Both Macready and Dayton models are thus statistically the most complete all-or-none models known at present. Nevertheless, two remarks should be made.

FORGETTING IMPLIES GUESSING

Statistical completeness is one requirement that models for mastery decisions should meet. Showing plausibly what happens when an examinee encounters an item is another. With regard to the latter, the plausibility of equation 1, the cornerstone of the Macready and Dayton models, is doubtful. If one bears in mind that both models were designed for items to which the right answer may be found by random

guessing, it is more realistic to assume that a master who, as a consequence of forgetting, happens not to know the answer to an item will do the best he or she can by guessing. There is not much difference between a nonmaster who does not "really" know the answer to an item and a master who has forgotten it. Both are in the same situation of not being able to produce the right answer and will react similarly. Adopting this knowledge or guessing standpoint for masters, the probability statement equation (1) should be replaced by

$$P_i(+) = \begin{cases} 1 - \beta_i' + \alpha_i' \beta_i' & \text{for a master} \\ \alpha_i' & \text{for a nonmaster,} \end{cases} \quad (3)$$

where for the sake of clearness the parameters have been apostrophized. From equation (3) a corrected version of the Macready and Dayton Model I can be derived as

$$P(j) = \left[\prod_{i=1}^n \alpha_i' a_{ij} (1 - \alpha_i')^{1-a_{ij}} \right] \bar{\theta} + \left[\prod_{i=1}^n \{\beta_i' (1 - \alpha_i')\}^{1-a_{ij}} \{1 - \beta_i' (1 - \alpha_i')\}^{a_{ij}} \right] \theta. \quad (4)$$

A correct version of Model II is derived accordingly.

Formally, the number of parameters in equations (2) and (4) are equal; only a reparameterization of the Macready and Dayton model has taken place with

$$\alpha_i' = \alpha_i \quad (5)$$

and

$$\beta_i' = \beta_i / (1 - \alpha_i). \quad (6)$$

equation (6) shows that Macready and Dayton's forgetting parameter has been given a new structure and interpretation: "forgetting the right answer" in the original model has been replaced by "forgetting and misguessing the right answer" in the revised model.

It should be noted that $\beta'_i \geq \beta_i$, which means that in the Macready and Dayton model the presence of forgetting is underestimated. The larger the value of α_i , the larger this underestimation. Macready and Dayton (1977) reported the results of an analysis based on Models I and II of four randomly selected items from each of two domains for 284 students. Their estimated alpha and beta values vary between .00 - .08 and .12 - .57, respectively. Using equation (6), the size of their underestimation of β'_i can be estimated. The estimated absolute bias varies between .00 - .07, while the estimated relative bias shows values between .00 - .28. Since the parameters in the revised model are one-to-one functions of the parameters in the original model, the invariant property of maximum likelihood estimators applies (Graybill, 1961, p. 36). Maximum likelihood estimates of the former can thus be obtained by maximum likelihood of the latter and, after that, by applying equations (5) and (6). Therefore, Macready and Dayton's computer programs MODEL 3 and MODEL 3G are, with a small adjustment, still usable for estimation purposes.

The correction for guessing introduced in equation (3) is based on a knowledge or random guessing model. The same model underlies the practice of formula scoring known from classical test theory. Formula scoring is exposed to the criticism that it does not take into account the partial information or misinformation an examinee may have about an item and that may influence his or her response to the item (Lord & Novick, 1968, pp. 303-310). This criticism does not apply here, because instead of using $1/A_i$, where A_i is the number of options of item i , the probability of guessing right α'_i is estimated from actual test data. For this reason α'_i is comparable with the guessing parameter in the three-parameter logistic model (see equation (7)).

A LATENT TRAIT APPROACH

Following Emrick and Adams (1969), Emrick (1971), and Besel (1973), Macready and Dayton (1977) conceived mastery as a latent class, and their two models are an application of latent class analysis to mastery testing. Alternatively, an application of latent trait theory to mastery testing can be adopted. In this application mastery is conceived as a region on a latent variable underlying a homogeneous, objective-based item domain and representing the level of competence an examinee may have with respect to this domain. Approaches to assessing mastery in which an underlying, continuous variable is assumed are given by Birnbaum (1968, chapter 19); Hambleton and Novick (1973); Huynh (1976); Mellenbergh, Koppelaar, and van der Linden (1977); Millman (1973); Novick and Lewis (1974); van der Linden and Mellenbergh (1977); and Wilcox (1976).

A latent trait model increasingly applied to testing problems is the three-parameter logistic model that gives the probability of a successful response to item i as

$$P_i(+) = c_i + (1-c_i) \Psi \left[a_i (\theta - b_i) \right], \quad (7)$$

where

- θ is the latent trait,
- a_i is the discriminating power of item i ,
- b_i is the difficulty of item i ,
- c_i is the probability of guessing item i right, and
- Ψ is the logistic distribution function

(Birnbaum, 1968, pp. 399-405). In order to consider the revised model equation (4) from a latent trait theoretic standpoint, note how equation (3) can be rewritten as

$$P_i(+) = \begin{cases} \alpha_i' + (1-\alpha_i')(1-\beta_i') \\ \alpha_i' \end{cases} \quad (8)$$

and that α'_i has the same position and meaning as the parameter c_i in the above logistic model. Adopting a latent cutting score θ_c such that examinees with $\theta \geq \theta_c$ are deemed to be a master (see Huynh, 1976), equation (8) can be considered a latent trait model obtained from equation (7) by

$$\Psi \left[a_i (\theta - b_i) \right] := \begin{cases} 1 - \beta'_i & \text{for } \theta \geq \theta_c \\ 0 & \text{for } \theta < \theta_c \end{cases} \quad (9)$$

Thus, according to a latent trait theoretic point of view, the revised Macready and Dayton models require item characteristic functions with a form as displayed in Figure 1: step functions with a single step just above θ_c .

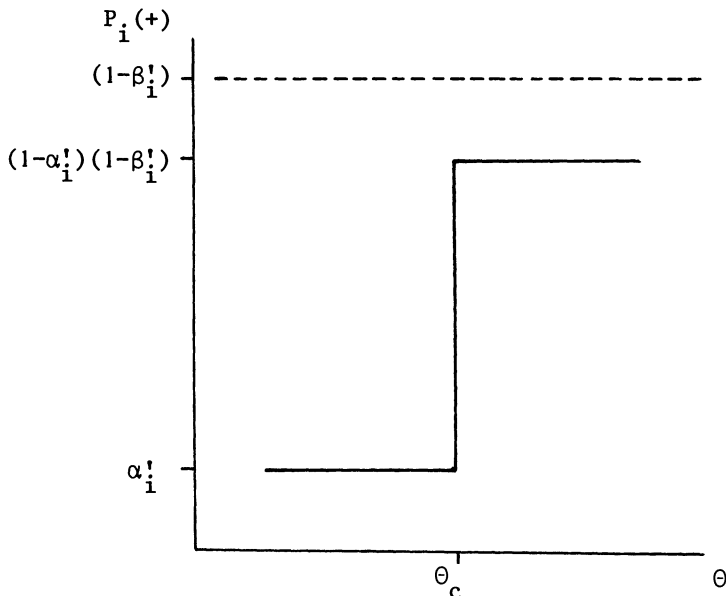


FIGURE 1

Item Characteristic Curve Required for the Revisited
Macready and Dayton Models

Note that for all values of θ , however small or large they are, the probability of a successful response to item i is equal to either α_i' or $(1-\alpha_i')(1-\beta_i')$. For items in which no guessing is possible $\alpha_i' = 0$, and Figure 1 takes the form of the right-hand part of substitution equation (9); the larger value equals $1-\beta_i'$ and the lower value zero.

As indicated above, the parameter α_i' has the same position and meaning as the guessing parameter c_i in the Birnbaum model. It is also interesting to interpret the forgetting parameter β_i' from a latent trait standpoint. According to latent trait theory, β_i' may be considered equal to

$$1 - \left[\int_{\theta \geq \theta_c} P_i(+)|g(\theta)|d\theta \right] / \left[\int_{\theta \geq \theta_c} g(\theta)d\theta \right], \quad (10)$$

where $g(\theta)$ is the probability density of θ , and is simply to be interpreted as the probability that a random master does not know the right answer to item i .

MASTERY: LATENT CLASS OR REGION ON A LATENT VARIABLE?

Which model is more plausible: the Macready and Dayton model conceiving mastery as a latent class or the latent trait model equation (7) to which a cutting score is added defining mastery as a region on a latent variable? In answering this question, it should be realized that the two models are incompatible. From the latent trait standpoint, the latent class model implies an unrealistic item characteristic curve, whereas from the latent class standpoint it is unrealistic to assume an item characteristic curve. In the opinion of the author there is much that argues against the latent class and for the latent trait conception. Experience shows that classroom learning is often a process in which knowledge is gathered gradually. Moreover, the results of most learning experiments can best be approximated by learning curves that express gradual accumulation of knowledge. Nevertheless, the Macready and Dayton model

assumes that knowledge has an all-or-none character and that learning is a process in which the student at a certain moment jumps from the state of not knowing any item to the state of knowing perfectly all items of the domain. Similarly, the adoption of a true score vector that contains only "ones" for a master and "zeros" for a nonmaster seems to be unrealistic for the kind and size of item domains usually constructed for mastery learning programs. In most instances it is incorrect to assume for such domains that $\bar{\theta} \times 100\%$ of the examinees does not know any item and guesses at all items, whereas the remaining percentage of examinees knows all items, but happens to forget the right answer to some of them, even if the model is merely meant as an idealistic approximation of the true state of affairs.

Theoretically, one type of item domain is possible for which the learning process and score vectors implied by the Macready and Dayton model seem to be meaningful: a domain in which all items have Guttman characteristic curves with their jump at θ_c . Bearing in mind that $P_i(+)$ from latent trait theory is identical to the true item score in the sense of classical test theory, it is obvious that for this type of domain all examinees have a true score vector with "zeros" and "ones", and that passing θ_c during the learning process gives rise to learning curves showing a single step. Although Guttman items at θ_c would be optimal for mastery testing (van der Linden, 1978), it is unrealistic to assume that this optimum will ever be reached. It is rare to find an item with a value for its discriminating power parameter a_i as large as two (Lord & Novick, 1968, p. 379). Furthermore, constructing a sufficiently large domain with items all having the value of θ_c for their difficulty parameter b_i may be considered practically impossible. It is therefore inadequate to assume the fulfillment of the condition of a domain containing only Guttman items at θ_c as a starting point for modeling what happens when a test with items from the domain is administered to examinees. A better strategy seems to be to adopt a latent trait conception and to base analyses of mastery testing on a model like the three-parameter logistic model equation(7). The item characteristic curve of this model is more flexible and general than the item characteristic curve implied by the Macready and Dayton model. In addition to a guessing parameter, it contains parameters for the discriminating power and the difficulty

of the item. Therefore, whenever it happens that item domains are met for which the Macready and Dayton model gives a reasonable good approximation, analyses based on the three-parameter logistic model will apply as well. For an introduction to two-point classification problems that are based on this model and of importance to analyses of mastery testing, the reader is referred to Birnbaum (1968, chapter 19).

It should be noted that Figure 1 does not display a chance-corrected version of a Guttman item characteristic curve; its upper limit has the value of $1 - \beta'_i$ instead of one.

Considering the Macready and Dayton model from a latent trait standpoint, this means that beyond Θ_c an increase in competence does not produce an increase in the probability of a successful response to the item. No examinee, however competent, will respond with almost sure success to the item. Note also that in the Macready and Dayton model the forgetting parameter β'_i is a function of the item. Forgetting is considered a process depending on the item; examinees with $\Theta \geq \Theta_c$ will forget the answer to the item with a probability determined by the item, no matter how long they have studied and how much time has elapsed since their last learning. Accordingly, for modern learning strategies like mastery learning, it is better to assume, for learning objectives and item domains normally encountered in educational practice, that examinees can improve their results by studying further under optimal learning conditions. In contrast to the Macready and Dayton model, the three-parameter logistic model equation (7) is consistent with this point of view: the probability of a successful response is considered a function of the competence level Θ , and by enhancing competence, this probability can be increased until a successful response is certain.

The question may be raised whether the parameter β_i in the Macready and Dayton model is indeed to be interpreted as a forgetting parameter. Formally, β_i is the probability that an examinee who is classified as a master is not able to produce the right answer. Interpreting β_i as a forgetting parameter suggests that the examinees have attained the status of master, but that at the moment of testing, for instance, because of time elapse, the probability of a successful response to the item is smaller than one. From a latent trait

standpoint, however, it is clear that β_i may be written as in equation (10), and is simply the possibility that a student randomly drawn from the subpopulation of masters does not know the right answer to item i . Note how this probability is written as a function of the actual competence level of all masters at the moment of testing. No idealistic state of mastery is assumed, and neither are "earthly" disturbances like forgetting, which prevent "true" masters from expressing their state in responding to the item, assumed. Note also how the interpretation of β_i as a forgetting parameter supports the idealistic all-or-none conception of mastery. Items for which masters at the moment of testing have a true score or, what is the same, a probability of success less than one are simply dealt with as items that elicit forgetting. By assuming an item parameter β_i for forgetting, the items are blamed for what--from a latent trait standpoint--seems to be an inadequate conception of mastery.

ACKNOWLEDGMENTS

The author thanks Fred N. Kerlinger and Gideon J. Mellenbergh for their comments.

REFERENCES

- Besel, R. Using group performance to interpret individual responses to criterion-referenced tests. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, Louisiana, February-March 1973 (EDRS No. ED 076 658).
- Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley, 1968.
- Emrick, J.A. An evaluation model for mastery learning. Journal of Educational Measurement, 1971, 8, 321-326.

- Emrick, J.A., & Adams, E.N. An evaluation model for individualized instruction (Report RC2674). Yorktown Heights, New York: IBM Thomas J. Watson Research Center, October 1969.
- Graybill, F.A. An introduction to linear statistical models. New York: McGraw-Hill, 1961.
- Hambleton, R.K., & Novick, M.R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.
- Huynh, H. Statistical considerations of mastery scores. Psychometrika, 1976, 41, 65-79.
- Lord, F.M. & Novick, M.R. Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley, 1968.
- Macready, G.B., & Dayton, C.M. The use of probabilistic models in the assessment of mastery. Journal of Educational Statistics, 1977, 2, 99-120.
- Mellenbergh, G.J.; Koppelaar, H.; & van der Linden, W.J. Dichotomous decisions based on dichotomously scored items: a case study. Statistica Neerlandica, 1977, 31, 161-169.
- Meskauskas, J.A. Evaluation models for criterion-referenced testing: views regarding mastery and standard-setting. Review of Educational Research, 1976, 46, 133-158.
- Millman, J. Passing scores and test lengths for domain-referenced measurement. Review of Educational Research, 1973, 43, 205-216.
- Novick, M.R., & Lewis, C. Prescribing test length for criterion-referenced measurement. In C.W. Harris, M.C. Alkin, & W.J. Popham (Eds.), Problems in criterion-referenced measurement. (CSE monograph series in evaluation No. 3.) Los Angeles: Center for the Study of Evaluation, University of California, 1974.
- van der Linden, W.J. A latent trait look at pretest-post-test validation of criterion-referenced items (Twente Educational Report No. 7). Enschede, The Netherlands: Twente University of Technology, Department of Applied

Education, 1978.

van der Linden, W.J., & Mellenbergh, G.J. Optimal cutting scores using a linear loss function. Applied Psychological Measurement, 1977, 1, 593-599.

Wilcox, A. A note on the length and passing score of a mastery test. Journal of Educational Statistics, 1976, 1, 359-364.

AUTHOR

VAN DER LINDEN, WIM J. Address: Vakgroep Onderwijskunde, Technische Hogeschool Twente, Postbus 217, 7500 AE Enschede, The Netherlands. Title: Wetenschappelijk medewerker. Degrees: Doctoraal examen, Rijksuniversiteit Utrecht. Specialization: psychometric theory and research methodology.